Bayesian Inversion Applied to an Ice Sheet Flow Problem and to Power Grid

Noémi Petra

joint work with:
Omar Ghattas, Toby Isaac, James Martin, and Georg Stadler
Mihai Anitescu, Emil Constantinescu, Cosmin Petra, and Zhang Zheng

1 School of Natural Sciences, University of California, Merced
2 ICES, The University of Texas at Austin
3 Courant Institute of Mathematical Sciences, New York University
4 Argonne National Laboratory
5 Lawrence Livermore National Laboratory

Optimization Opening Workshop
Statistical and Applied Mathematical Sciences Institute (samsi)
August 30, 2016
Outline

1. Ice sheet flow problem
   - Forward and inverse ice sheet flow problems
   - Adjoint-based solution method: derivatives computation

2. Bayesian Inverse Problems
   - Bayesian approach to inverse problems
   - Challenges for large-scale Bayesian inverse problems

3. Approximations of the posterior pdf
   - Gaussian approximation around MAP
   - (Stochastic Newton) MCMC sampling
   - Inference/uncertainty quantification applied to ice sheet models

4. Dynamic parameter estimation: inertia estimation in electric power systems
   - Motivation and background
   - The power grid inertia estimation problem: the MAP point
   - Gaussian approximation of the posterior
Outline

1 Ice sheet flow problem
   - Forward and inverse ice sheet flow problems
   - Adjoint-based solution method: derivatives computation

2 Bayesian Inverse Problems
   - Bayesian approach to inverse problems
   - Challenges for large-scale Bayesian inverse problems

3 Approximations of the posterior pdf
   - Gaussian approximation around MAP
   - (Stochastic Newton) MCMC sampling
   - Inference/uncertainty quantification applied to ice sheet models

4 Dynamic parameter estimation: inertia estimation in electric power systems
   - Motivation and background
   - The power grid inertia estimation problem: the MAP point
   - Gaussian approximation of the posterior
Balance of linear momentum, mass, and energy

\[-\nabla \cdot \left[ \eta(\theta, u) \dot{\varepsilon} - I_p \right] = \rho g, \quad [\dot{\varepsilon} = \frac{1}{2} (\nabla u + \nabla u^T)]\]

\[\nabla \cdot u = 0,\]

\[\rho c \left( \frac{\partial \theta}{\partial t} + u \cdot \nabla \theta \right) - \nabla \cdot (K \nabla \theta) = 2 \eta \text{tr}(\dot{\varepsilon}^2)\]

Mathematical challenges:
- highly ill-conditioned linear systems
- complex, high aspect ratio geometry
- strong nonlinearities
- multiphysics couplings
- uncertain basal boundary conditions, topography, heat flux
- observational data: InSAR, laser altimetry, GRACE satellite, ice cores, radar

Inference of basal bdry. cond.
- critical for climate simulations
- inference of effective sliding/friction coefficient $\beta$ in Robin boundary condition

\[T(\sigma n) + \beta(x) Tu = 0\] (tangential component) from surface velocity observations
Data $\rightarrow$ Inference $\rightarrow$ Prediction for Ice Sheets

(The uncertainty in the projections of the land ice contributions (to sea level rise) is dominated by the various uncertainties in the land ice models themselves ... rather than in the temperature projections.)” (From: IPCC Climate Change Report 2007)

Ultimate target:

Data: Current ice sheet geometry and surface ice flow velocity

Inference: For unknown basal boundary conditions and current state of the ice sheet under certain stationarity assumptions

Prediction: Sea level change by 2100 or 2200 under different climate scenarios, ideally, in a coupled atmosphere/ocean/ice climate model
Data $\rightarrow$ Inference $\rightarrow$ Prediction for Ice Sheets

(T)he uncertainty in the projections of the land ice contributions (to sea level rise) is dominated by the various uncertainties in the land ice models themselves ... rather than in the temperature projections.’’ (From: IPCC Climate Change Report 2007)

**Ultimate target:**

**Data:** Current ice sheet geometry and surface ice flow velocity

**Inference:** For unknown basal boundary conditions and current state of the ice sheet under certain stationarity assumptions

**Prediction:** Sea level change by 2100 or 2200 under different climate scenarios, ideally, in a coupled atmosphere/ocean/ice climate model

**In this presentation:**

- Full nonlinear Stokes ice sheet model
- InSAR Antarctic surface velocity data
- Bayesian inference assuming additive Gaussian noise and Gaussian prior
- Gaussian approximation for basal boundary condition field distribution
- Simplified prediction (current ice sheet mass loss)
- Linearized parameter-to-prediction map

Noemi Petra, UCM

Bayesian Inverse Problems

August 30, 2016
Nonlinear Stokes ice sheet model

The flow of ice is commonly modeled as a viscous, shear-thinning, incompressible fluid via the balance of mass and linear momentum:

\[
- \nabla \cdot [2\eta(u, n) \dot{\varepsilon}_u - Ip] = \rho \mathbf{g} \quad \text{in } \Omega \\
\nabla \cdot \mathbf{u} = 0 \quad \text{in } \Omega \\
\sigma_u n = 0 \quad \text{on } \Gamma_t \\
u \cdot n = 0, \ T\sigma_u n + \exp(\beta) Tu = 0 \quad \text{on } \Gamma_b
\]

- \(\mathbf{u}\) ice flow velocity, \(p\) pressure
- \(\sigma_u = -Ip + 2\eta(u, n)\dot{\varepsilon}_u\) stress tensor
- \(\dot{\varepsilon}_u = \frac{1}{2}(\nabla u + \nabla u^T)\) strain rate tensor
- \(\eta(u, n) = \frac{1}{2}A^{-\frac{1}{n}}\dot{\varepsilon}_u^\frac{1-n}{2n}\) effective viscosity
- \(\dot{\varepsilon}_\Pi = \frac{1}{2}\text{tr}(\dot{\varepsilon}_u^2)\) second invariant of the strain rate tensor
- \(\rho\) density, \(\mathbf{g}\) gravity
- \(\mathbf{n}\) unit normal vector
- \(\beta\) log basal sliding coefficient
- \(T = I - n \otimes n\) tangential operator
- \(\Gamma_t\) and \(\Gamma_b\) top and base boundaries
InSAR observed surface ice velocity dataset

- Assembled from multiple satellite interferometric synthetic-aperture radar data from 2007--2009
- Data set integrates 900 satellite tracks and more than 3,000 orbits
Inversion for the basal friction coefficient

- Regularized data misfit cost functional:

\[
\min_{\beta} J(\beta) := \frac{1}{2} \int_{\Gamma_t} (B u(\beta) - d)^2 \, ds + \frac{\gamma}{2} \int_{\Gamma_b} \nabla_{\Gamma} \beta \cdot \nabla_{\Gamma} \beta \, ds
\]

- Inexact Newton: solve by CG inexactley and operator-free,

\[
\mathcal{H}(\beta) \tilde{\beta} = -\mathcal{G}(\beta), \quad \beta^{\text{new}} = \beta + \alpha \tilde{\beta}
\]

- Gradient given by:

\[
\mathcal{G}(\beta) := \exp(\beta) \, T u \cdot T v + \nabla_{\Gamma} \cdot (\gamma \nabla_{\Gamma} \beta)
\]

- \(u\) state velocity, \(v\) adjoint velocity, \(\beta\) basal friction field
- \(d\) observed surface velocity, \(B\) observation operator
- \(R\) regularization term, \(\mathcal{H}\) Hessian operator

Gradient computation: additional adjoint Stokes eqn

- \( u \) and \( p \) satisfy the forward (nonlinear) Stokes equations
  \[
  \nabla \cdot u = 0 \quad \text{in} \ \Omega
  \]
  \[
  -\nabla \cdot \left[ \eta(u)(\nabla u + \nabla u^T) - Ip \right] = \rho g \quad \text{in} \ \Omega
  \]
  \[
  \sigma_u n = 0 \quad \text{on} \ \Gamma_t
  \]
  \[
  u \cdot n = 0, \ T\sigma_u n + \exp(\beta) Tu = 0 \quad \text{on} \ \Gamma_b
  \]

- \( v \) and \( q \) satisfy the adjoint Stokes equations
  \[
  \nabla \cdot v = 0 \quad \text{in} \ \Omega
  \]
  \[
  -\nabla \cdot \sigma_v = 0 \quad \text{in} \ \Omega
  \]
  \[
  \sigma_v n = -B^*(Bu - d) \quad \text{on} \ \Gamma_t
  \]
  \[
  v \cdot n = 0, \ T\sigma_v n + \exp(\beta) Tv = 0 \quad \text{on} \ \Gamma_b
  \]

where the adjoint stress \( \sigma_v \) is
\[
\sigma_v := 2\eta(u) \left( I + \frac{1}{n} \frac{\dot{\varepsilon}_u \otimes \dot{\varepsilon}_u}{\dot{\varepsilon}_u \cdot \dot{\varepsilon}_u} \right) \dot{\varepsilon}_v - Iq
\]
Hessian action: 2 additional (linearized) Stokes-like eqns

- Action of Hessian operator in direction $\hat{\beta}$ evaluated at $\beta$

$$\mathcal{H}(\beta)\hat{\beta} := \exp(\beta)(\hat{\beta}Tu \cdot Tv + T\hat{u} \cdot Tv + Tu \cdot T\hat{v}) + \nabla \Gamma \cdot (\gamma \nabla \Gamma \hat{\beta})$$

- where $\hat{u}$ and $\hat{p}$ satisfy the incremental forward equations

$$\nabla \cdot \hat{u} = 0 \quad \text{in} \; \Omega$$
$$-\nabla \cdot \sigma_{\hat{u}} = 0 \quad \text{in} \; \Omega$$
$$\sigma_{\hat{u}} n = 0 \quad \text{on} \; \Gamma_t$$
$$\hat{u} \cdot n = 0, \quad T\sigma_{\hat{u}} n + \exp(\beta)T\hat{u} = -\hat{\beta} \exp(\beta)Tu \quad \text{on} \; \Gamma_b$$

with $\sigma_{\hat{u}} := 2\eta(u) \left( 1 + \frac{1-n}{n} \frac{\dot{e}_u \otimes \dot{e}_u}{\dot{e}_u \cdot \dot{e}_u} \right) \dot{e}_u - I\hat{p}$
Hessian action: 2 additional (linearized) Stokes-like eqns

- Action of Hessian operator in direction \( \hat{\beta} \) evaluated at \( \beta \)
  \[
  \mathcal{H}(\beta)\hat{\beta} := \exp(\beta)(\hat{\beta}Tu \cdot Tv + T\hat{u} \cdot Tv + Tu \cdot T\hat{v}) + \nabla_{\Gamma} \cdot (\gamma \nabla_{\Gamma} \hat{\beta})
  \]

- where \( \tilde{v}, \tilde{q} \) satisfy the incremental adjoint equations
  \[
  \nabla \cdot \hat{v} = 0 \quad \text{in } \Omega \\
  -\nabla \cdot \sigma_{\hat{v}} = -\nabla \cdot \tau_{\hat{u}} \quad \text{in } \Omega \\
  \sigma_{\hat{v}} n = -B^*B\hat{u} - \tau_{\hat{u}} n \quad \text{on } \Gamma_t \\
  \hat{v} \cdot n = 0, \quad T\sigma_{\hat{v}} n + \exp(\beta)T\hat{v} = -T\tau_{\hat{u}} n \quad \text{on } \Gamma_b
  \]

with \( \sigma_{\hat{v}} := 2\eta(u) \left(1 + \frac{1-n}{n} \frac{\dot{u} \otimes \dot{u}}{\dot{u} \cdot \dot{u}}\right) \dot{\hat{v}} - I\dot{q}, \) and \( \tau_{\hat{u}} = 2\eta(u)\Psi \dot{\hat{u}}, \) where

\[
\Psi = (1+\frac{1}{n} \frac{\dot{u} \cdot \dot{u}}{n}) l + \frac{1-n}{n} \left[ \frac{\dot{u} \otimes \dot{u}}{\dot{u} \cdot \dot{u}} + 2 \frac{\dot{u} \otimes \dot{v}}{\dot{u} \cdot \dot{u}} + \frac{1-3n}{n} \frac{\dot{u} \otimes \dot{u}}{(\dot{u} \cdot \dot{u})^2} \right]
\]

Noemi Petra, UCM

Bayesian Inverse Problems

August 30, 2016 10 / 44
Summary of the components of inexact Newton-CG

- The Newton system is solved inexactely by early termination of CG iterations via Eisenstat--Walker (to prevent oversolving) and Steihaug (to avoid negative curvature) criteria.
- Gradients and Hessian actions at each CG iteration are expressed in terms of solutions of forward and adjoint PDEs, and their linearizations.
- Preconditioning is effected by the inverse of the (elliptic) regularization operator, which is carried out by multigrid V-cycles.
- Globalization is by an Armijo backtracking line search.
- Continuation on the regularization is carried out to warm-start the Newton iterations, i.e., we initially use a large value of $\gamma$ and decrease it during the iteration to the desired value.
- Parallel implementation of all components, whose cost is dominated by solution of forward and adjoint PDEs and evaluation of inner product-like quantities to construct gradient and Hessian action quantities.
ymir*: forward ice sheet flow solver

- **accuracy**: models state equations in 3D, finite element velocity-discontinuous pressure pair $\mathbb{Q}_p \times \mathbb{Q}^{\text{disc}}_{p-2}$ discretization for element-wise mass conservation, mesh adaptivity allows resolving flow features at appropriate length scales;

- **efficiency**: parallel, highly scalable octree-based AMR provided by $\texttt{p4est}$ library, Newton-Krylov method to solve nonlinear equations, scalable solvers for Newton step;

- **robustness**: parameter regimes may vary widely during optimization.

*Toby Isaac, PhD Thesis, ICES UT Austin (advisors: Omar Ghattas and Georg Stadler)
Antarctic inversion for $\beta$: InSAR data

Setup & performance

- Spatial discretization (all on same mesh):
  - velocity (state, adjoint, incremental state, incremental adjoint): $Q_2$
  - pressure (state, adjoint, incremental state, incremental adjoint): $Q_{\text{disc}}$
  - basal friction, incremental basal friction: $Q_2$
    (biquadratic)

- # state parameters: 4,085,841
- # inversion parameters: 409,545
- # elements: 99,984
- # of cores: 1024
- inexact Newton-CG
- use PETSc’s framework to solve the linear systems of equations
- real data
- reduction in norm of gradient: $10^{-3}$
- # of Newton iterations: 213
- average # of CG iterations per Newton iteration: 239
- total # of (linearized) Stokes: 107,578

Stampede, Texas Advanced Computing Center (TACC)
Antarctic ice sheet inversion for basal friction field: InSAR data

Left: InSAR-based Antarctica ice surface velocity observations
Right: Inferred basal friction field
Antarctic ice sheet inversion for basal friction field:
InSAR data

Left: Recovered ice surface velocity observations
Right: Inferred basal friction field
Scalability of the inverse solver

Inexact Newton-CG method applied to an ice sheet inverse problem

<table>
<thead>
<tr>
<th>#sdof</th>
<th>#pdof</th>
<th>#N</th>
<th>#CG</th>
<th>avgCG</th>
<th>#Stokes</th>
</tr>
</thead>
<tbody>
<tr>
<td>95,796</td>
<td>10,371</td>
<td>42</td>
<td>2718</td>
<td>65</td>
<td>7031</td>
</tr>
<tr>
<td>233,834</td>
<td>25,295</td>
<td>39</td>
<td>2342</td>
<td>60</td>
<td>6440</td>
</tr>
<tr>
<td>848,850</td>
<td>91,787</td>
<td>39</td>
<td>2577</td>
<td>66</td>
<td>6856</td>
</tr>
<tr>
<td>3,372,707</td>
<td>364,649</td>
<td>39</td>
<td>2211</td>
<td>57</td>
<td>6193</td>
</tr>
<tr>
<td>22,570,303</td>
<td>1,456,225</td>
<td>40</td>
<td>1923</td>
<td>48</td>
<td>5376</td>
</tr>
</tbody>
</table>

- **#sdof**: number of degrees of freedom for the state variables;
- **#pdof**: number of degrees of freedom for the inversion parameter field;
- **#N**: number of Newton iterations;
- **#CG, avgCG**: total and average (per Newton iteration) number of CG iterations;
- **#Stokes**: total number of linear(ized) Stokes solves (from forward, adjoint, and incremental forward and adjoint problems)

The cost of solving the inverse problem by the inexact Newton-CG method, measured by the number of Stokes solves, is independent of the parameter dimension as well as the data dimension.
Outline

1. **Ice sheet flow problem**
   - Forward and inverse ice sheet flow problems
   - Adjoint-based solution method: derivatives computation

2. **Bayesian Inverse Problems**
   - Bayesian approach to inverse problems
   - Challenges for large-scale Bayesian inverse problems

3. **Approximations of the posterior pdf**
   - Gaussian approximation around MAP
   - (Stochastic Newton) MCMC sampling
   - Inference/uncertainty quantification applied to ice sheet models

4. **Dynamic parameter estimation: inertia estimation in electric power systems**
   - Motivation and background
   - The power grid inertia estimation problem: the MAP point
   - Gaussian approximation of the posterior
Bayesian approach to inverse problems

Inverse problem

\[ d = f(m) + e \]

Interpret \( m, d \) as random variables; solution of inverse problem is a probability density function \( \pi_{\text{post}}(m) \) for \( m \):

Remarks:

- systematic method to quantify measurement errors and incorporate prior knowledge
- allows quantification of uncertainty in reconstruction
- related to regularization approach (think of \( \pi_{\text{post}}(m) \) as \( \exp(-J(m)) \))
- high-dimensional probability density
Bayesian approach to inverse problems

Inverse problem

\[ d = f(m) + e \]

Interpret \( m, d \) as random variables; solution of inverse problem is a probability density function \( \pi_{\text{post}}(m) \) for \( m \):

Target:
- characterize \( \pi_{\text{post}}(m) \) statistically (mean, covariance, MAP point. . . )
- for functions \( m \) (large vectors after discretization)
- for expensive \( f(\cdot) \)
- exploit connection to PDE-constrained optimization
Bayes formula (finite dimensions)

Given:

\[ \pi_{pr}(m) : \text{prior p.d.f. of model parameters } m \]
\[ \pi_{obs}(d) : \text{prior p.d.f. of measurement error } d \]
\[ \pi_{model}(d|m) : \text{conditional p.d.f. combining } d \text{ and } m \text{ (model)} \]

Then, the posterior p.d.f. of the model parameters is given by:

\[ \pi_{post}(m|d) \propto \pi_{pr}(m) \pi_{like}(d|m) \]
Bayes formula (infinite dimensions)

(A. Stuart, Acta Numerica, (2010))

- invert for $m \in L^2(\Omega)$
- Gaussian random field prior, $\mu_0 := \mathcal{N}(m_0, A^{-2})$ on $L^2(\Omega)$
- covariance operator is given by the inverse of differential operator $A^2$
  where $A(m) := -\alpha \nabla \cdot (\Theta \nabla m) + \alpha m$
- Bayesian solution of the inverse problem defined as conditional measure $\mu^d$ of $m$ given the data $d \in \mathbb{R}^q$, where

$$
\frac{d\mu^d}{d\mu_0} = \frac{1}{Z_{\text{like}}(d|m)} \propto \exp \left( -\frac{1}{2} \| f(m) - d \|^2_{\Gamma_{\text{noise}}^{-1}} \right)
$$

is the Radon-Nikodym derivative w.r.t. $\mu_0$ and $f(m)$ is the parameter-to-observable map and $\Gamma_{\text{noise}}$ the noise covariance operator

- leads to well-posed Bayesian inverse problem (in 2D and 3D); exploits fast elliptic solvers
**Gaussian prior and noise**

We assume additive Gaussian noise in the measurements:

\[ d = f(m) + e, \quad e \sim \mathcal{N}(0, \Gamma_{\text{noise}}) \]

Thus:

\[ \pi_{\text{like}}(d|m) = \exp\left(-\frac{1}{2}(f(m) - d)^T \Gamma_{\text{noise}}^{-1} (f(m) - d)\right) \]

If the prior is Gaussian with mean \( m_{\text{prior}} \) and covariance \( \Gamma_{\text{prior}} \), then we obtain for the posterior pdf:

\[ \pi_{\text{post}}(m) \propto \exp\left(-\frac{1}{2} \| f(m) - d \|_{\Gamma_{\text{noise}}^{-1}}^2 - \frac{1}{2} \| m - m_{\text{prior}} \|_{\Gamma_{\text{prior}}^{-1}}^2 \right) \]

The ``maximum a posteriori'' point is

\[ m_{\text{MAP}} \overset{\text{def}}{=} \arg \max_m \pi_{\text{post}}(m) \]

\[ = \arg \min_m \frac{1}{2} \| f(m) - d \|_{\Gamma_{\text{noise}}^{-1}}^2 + \frac{1}{2} \| m - m_{\text{prior}} \|_{\Gamma_{\text{prior}}^{-1}}^2 . \]

\( \Rightarrow \) deterministic inverse problem with appropriate weighted norms!
Challenges for large-scale Bayesian inverse problems

- Method of choice is to sample the posterior density using Markov chain Monte Carlo (MCMC).
- For inverse problems characterized by high-dimensional parameter spaces and expensive forward simulations, standard MCMC methods become prohibitive.
- Standard MCMC methods view the parameter-to-observable map as a black-box.

Goals:
- overcome bottlenecks of MCMC: develop MCMC methods that reduce effective problem dimension by exploiting the structure of the problem (as has been done successfully in PDE-constrained optimization)
- apply to large-scale Bayesian inverse problems
- approximate the posterior pdf
Outline

1. Ice sheet flow problem
   - Forward and inverse ice sheet flow problems
   - Adjoint-based solution method: derivatives computation

2. Bayesian Inverse Problems
   - Bayesian approach to inverse problems
   - Challenges for large-scale Bayesian inverse problems

3. Approximations of the posterior pdf
   - Gaussian approximation around MAP
   - (Stochastic Newton) MCMC sampling
   - Inference/uncertainty quantification applied to ice sheet models

4. Dynamic parameter estimation: inertia estimation in electric power systems
   - Motivation and background
   - The power grid inertia estimation problem: the MAP point
   - Gaussian approximation of the posterior
Approximations of the posterior pdf

With these assumptions (Gaussian prior and noise), the posterior is:

\[ \pi_{\text{post}}(m) \propto \exp \left( -\frac{1}{2} \| f(m) - d \|_{\Gamma_{\text{noise}}}^{-1} - \frac{1}{2} \| m - m_{\text{prior}} \|_{\Gamma_{\text{prior}}}^{-1} \right) \]

MAP estimation

Gaussian approximation around MAP

(Stochastic Newton) MCMC sampling
Gaussian approximation around MAP

Assume linear(ized) parameter-to-observable map

\[ d = Fm. \]

Then, the posterior p.d.f. is:

\[
\pi_{\text{post}}(m) \propto \exp \left( -\frac{1}{2} \left\| Fm - d \right\|^2_{\Gamma_{\text{noise}}^{-1}} - \frac{1}{2} \left\| m - m_{\text{prior}} \right\|^2_{\Gamma_{\text{prior}}^{-1}} \right)
\]

\[
= \exp \left( -\frac{1}{2} (m - m_{\text{MAP}})^T (F^T \Gamma_{\text{noise}}^{-1} F + \Gamma_{\text{prior}}^{-1}) (m - m_{\text{MAP}}) \right)
\]

Thus, the posterior is also Gaussian, i.e., \( m \sim \mathcal{N}(m_{\text{MAP}}, \Gamma_{\text{post}}) \). The covariance matrix is the inverse Hessian of \( J(\cdot) \), i.e.,

\[
\Gamma_{\text{post}}^{-1} = H = F^T \Gamma_{\text{noise}}^{-1} F + \Gamma_{\text{prior}}^{-1}
\]

\[
= \nabla^2_m (-\log \pi_{\text{post}})
\]
Low-rank-based posterior covariance

- Posterior covariance is given by prior covariance less information gained from data:

\[ \Gamma_{\text{post}} \approx \Gamma_{\text{prior}} - W_r D_r W_r^T \]

\( W_r \) contains only \( r \) (generalized) eigenvectors of the Hessian of the data misfit term, orthogonal with respect to \( \Gamma_{\text{prior}}^{-1} \).

- The spectrum of the prior-preconditioned data misfit Hessian:

Cost of low rank approximation (extracted via randomized SVD) measured in the number of matrix-free Hessian vector products is independent of the data and parameter dimensions.
Samples from prior (top) and posterior (bottom) pdfs

Data\,\rightarrow\text{Inference}\,\rightarrow\text{Prediction for Ice Sheets}

- Quantity of interest given by the ice mass flux to the ocean

\[ Q(\beta) := \int_{\Gamma_o} \rho \mathbf{u}(\beta) \cdot \mathbf{n} \, ds, \]

where \( \Gamma_o \) is an outflow boundary of interest.

- Linear approximation to parameter-to-prediction map

\[ Q_{\text{lin}}(\beta) = Q(\beta_{\text{MAP}}) + \langle g_\beta(\beta_{\text{MAP}}), \beta - \beta_{\text{MAP}} \rangle \]

\[ g_\beta(\cdot) := \frac{dQ(\cdot)}{d\beta} \] is the gradient with respect to \( \beta \)

- The linearized \( Q(\cdot) \) is distributed according to

\[ Q_{\text{lin}}(\cdot) \sim \mathcal{N}\left( Q(\beta_{\text{MAP}}), \langle g_\beta(\beta_{\text{MAP}}), \mathcal{H}^{-1}(\beta_{\text{MAP}})[g_\beta(\beta_{\text{MAP}})] \rangle \right) \]
The Jacobian of the parameter-to-prediction map

- The adjoint problem defined for the quantity $Q$:

\[
\begin{align*}
\nabla \cdot \mathbf{v} &= 0 \quad \text{in } \Omega \\
-\nabla \cdot \sigma_v &= 0 \quad \text{in } \Omega \\
\sigma_v \mathbf{n} &= 0 \quad \text{on } \Gamma_t \\
\sigma_v \mathbf{n} &= -\rho \mathbf{n} \quad \text{on } \Gamma_o \\
T\sigma_v \mathbf{n} + \exp(\beta_{\text{MAP}}) \mathbf{Tv} &= 0, \quad \mathbf{v} \cdot \mathbf{n} = 0 \quad \text{on } \Gamma_b
\end{align*}
\]

where the adjoint stress is given by

\[
\sigma_v := 2\eta(u) \left( 1 + \frac{1 - n \hat{\varepsilon}_u \otimes \hat{\varepsilon}_u}{n} \hat{\varepsilon}_u \cdot \hat{\varepsilon}_u \right) \hat{\varepsilon}_v - Iq.
\]

- The gradient of $Q$ at $\beta_{\text{MAP}}$:

\[
g_{\beta}(\beta_{\text{MAP}}) := \exp(\beta_{\text{MAP}}) \mathbf{Tu} \cdot \mathbf{Tv} \quad \text{on } \Gamma_b
\]
Inference for prediction

The gradient (left) and the ‘‘influential direction’’ in parameter space (right) for the ice mass flux from Totten Glacier to ocean. The mean and standard deviation of the prediction probability distribution for the ice mass flux is $71.24 \pm 0.30 \text{ Gt/a}$. 
(Stochastic Newton) MCMC sampling

Despite the explicit form of $\pi_{\text{post}}(m|d)$, its exploration is difficult due to:

- the high/infinite dimension of $m$
- the expensive PDE-based parameter-to-observable map $f$

Use sampling (Metropolis Hastings/Marcov chain Monte Carlo) to approximate statistics

- replace $\pi_{\text{post}}$ by a sample chain $m_k$
- sampling in high dimensions is challenging, requires many evaluations of $f$
Metropolis-Hastings Algorithm to sample the posterior

Random walk Metropolis

Choose initial parameters $m_0$, compute $\pi_{\text{post}}(m_0)$

for $k = 0, \ldots$ do

Draw sample $y$ from the proposal density $q(m_k, \cdot)$

Compute the acceptance probability:

$$\alpha(m_k, y) = \min \left\{ 1, \frac{\pi_{\text{post}}(y)q(y, m_k)}{\pi_{\text{post}}(m_k)q(m_k, y)} \right\}$$

Accept/Reject

end for

- common choice for the proposal density is an isotropic Gaussian (the resulting method is known as random walk Metropolis):

$$q(m_k, y) = \frac{1}{\sigma^n(2\pi)^{n/2}} \exp\left[ -\frac{(y - m_k)^T(y - m_k)}{2\sigma^2} \right]$$

- **Challenge/Goal**: devise a proposal density $q(m_k, y)$ that is both easy to sample and a good representation of the underlying posterior probability density.
Metropolis-Hastings Algorithm to sample the posterior

Stochastic Newton MCMC

Choose initial $m_0$, compute

\[ \pi_{\text{post}}(m_0), g(m_0) (\text{gradient}), H(m_0) (\text{Hessian}) \]

for $k = 0, \ldots$ do

Draw sample $y$ from the proposal density $q(m_k, \cdot)$

Compute the acceptance probability:

\[ \alpha(m_k, y) = \min \left\{ 1, \frac{\pi_{\text{post}}(y) q(y, m_k)}{\pi_{\text{post}}(m_k) q(m_k, y)} \right\} \]

Accept/Reject

end for

The contours of the stochastic Newton method proposal function overlayed on the Rosenbrock contours.

Proposal density for the stochastic Newton MCMC method:

\[ q(m_k, y) = \frac{\det H^{1/2}}{(2\pi)^{n/2}} \exp \left( -\frac{1}{2} (y - m_k + H^{-1} g)^T H (y - m_k + H^{-1} g) \right) \]

Modified stochastic Newton MCMC methods

- Stochastic Newton MCMC with dynamically changing Hessian (SN)
  \[
  q^{\text{SN}}(m_k, y) = \frac{\det H_k^{1/2}}{(2\pi)^{n/2}} \exp \left( -\frac{1}{2} \left( y - m_k + H_k^{-1}g_k \right)^T H_k \left( y - m_k + H_k^{-1}g_k \right) \right)
  \]

- Stochastic Newton MCMC method with MAP-based Hessian (SNMAP)
  \[
  q^{\text{SNMAP}}(m_k, y) \propto \exp \left( -\frac{1}{2} \left( y - m_k + H_{\text{MAP}}^{-1}g_k \right)^T H_{\text{MAP}} \left( y - m_k + H_{\text{MAP}}^{-1}g_k \right) \right)
  \]

- Independence sampling from the Gaussian approximation at the MAP point (SNIS)
  \[
  q^{\text{SNIS}}(m_{\text{MAP}}, y) \propto \exp \left( -\frac{1}{2} \left( y - m_{\text{MAP}} \right)^T H_{\text{MAP}} \left( y - m_{\text{MAP}} \right) \right)
  \]

Stochastic Newton MCMC (applied to Arolla glacier)

Performance results / Convergence diagnostics

<table>
<thead>
<tr>
<th>Method</th>
<th>MPSRF</th>
<th>IAT</th>
<th>ESS</th>
<th>MSJ</th>
<th>AAR</th>
<th>#Stokes</th>
<th>time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNIS</td>
<td>1.507</td>
<td>435</td>
<td>1207</td>
<td>280</td>
<td>9</td>
<td>4350</td>
<td>218</td>
</tr>
<tr>
<td>SNMAP</td>
<td>1.045</td>
<td>80</td>
<td>6563</td>
<td>190</td>
<td>6</td>
<td>960</td>
<td>48</td>
</tr>
<tr>
<td>SN</td>
<td>1.348</td>
<td>600</td>
<td>875</td>
<td>64</td>
<td>2</td>
<td>8400</td>
<td>420</td>
</tr>
</tbody>
</table>

- **MPSRF**: multivariate potential scale reduction factor
- **IAT**: integrated autocorrelation time
- **ESS**: effective sample size
- **MSJ**: mean squared jump distance
- **ARR**: average rejection rate
- **#Stokes**: # of Stokes solves per independent sample
- **time**: time per independent sample

**Statistics**:
- 21 parallel chains (each 25k); # samples: 525k; dof: 139;
- Rank of prior-preconditioned data misfit Hessian: 15

SNIS, SNMAP and SN are variants of the stochastic Newton MCMC method.

### Stochastic Newton MCMC (applied to Arolla glacier)

**Performance results / Convergence diagnostics**

<table>
<thead>
<tr>
<th></th>
<th>MPSRF</th>
<th>IAT</th>
<th>ESS</th>
<th>MSJ</th>
<th>AAR</th>
<th>#Stokes</th>
<th>time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNIS</td>
<td>1.507</td>
<td>435</td>
<td>1207</td>
<td>280</td>
<td>9</td>
<td>4350</td>
<td>218</td>
</tr>
<tr>
<td>SNMAP</td>
<td>1.045</td>
<td>80</td>
<td>6563</td>
<td>190</td>
<td>6</td>
<td>960</td>
<td>48</td>
</tr>
<tr>
<td>SN</td>
<td>1.348</td>
<td>600</td>
<td>875</td>
<td>64</td>
<td>2</td>
<td>8400</td>
<td>420</td>
</tr>
</tbody>
</table>

- **MPSRF**: multivariate potential scale reduction factor
- **IAT**: integrated autocorrelation time
- **ESS**: effective sample size
- **MSJ**: mean squared jump distance
- **ARR**: average rejection rate
- **#Stokes**: # of Stokes solves per independent sample
- **time**: time per independent sample

**Statistics**: 21 parallel chains (each 25k); # samples: 525k; dof: 139;

**Rank of prior-preconditioned data misfit Hessian**: 15

SNIS, SNMAP and SN are variants of the stochastic Newton MCMC method.

Posterior probability distribution

marginalized w.r.t. eigenvectors of the posterior covariance of the Gaussian app. at the MAP

- **Data-informed eigenvectors**: eigenvectors for which the information from the data dominates the information from prior (the variance in the posterior is significantly reduced due to the observations)
- **Shadowed eigenvectors**: eigenvectors for which the original prior variance was large, and yet the observations provide little information
- **Mixed eigenvectors**: eigenvectors for which the observations and the prior both have a significant influence
- **Prior-tail eigenvectors**: directions in which the prior is very certain
Joint Model and Parameter Dimension Reduction

To cope with the expensive high-dimensional state and parameter spaces:

1. identify a likelihood-informed parameter subspace that captures parameter directions where the change from prior to posterior is most significant and reduce the parameter dimension

2. accelerate the forward model evaluations (required when sampling) by identifying a low-dimensional subspace of the state and construct a reduced version of the forward model

3. combine 1 and 2 to accelerate the Bayesian solution of the inverse problem

Details in:
```````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````````
Outline

1 Ice sheet flow problem
   • Forward and inverse ice sheet flow problems
   • Adjoint-based solution method: derivatives computation

2 Bayesian Inverse Problems
   • Bayesian approach to inverse problems
   • Challenges for large-scale Bayesian inverse problems

3 Approximations of the posterior pdf
   • Gaussian approximation around MAP
   • (Stochastic Newton) MCMC sampling
   • Inference/uncertainty quantification applied to ice sheet models

4 Dynamic parameter estimation: inertia estimation in electric power systems
   • Motivation and background
   • The power grid inertia estimation problem: the MAP point
   • Gaussian approximation of the posterior
Dynamic parameter estimation in electric power systems

- State/parameter estimation is essential in operating the system: real-time monitoring and fault detection, dynamic/transient stability analysis, transmission switching, and many others.

- Controllers are combined with data acquisition devices:
  - SCADA (supervisory control and data acquisition) - 10 seconds sampling/response rate
  - PMU (phasor measurement unit) - 30 ms sampling rate, more accurate

- The goal is to build a dynamic estimation framework: given measurements and DAE model, infer the state/parameters of the power system.

- Also would like to characterize the uncertainties in the estimation.
Dynamic power grid simulation

The 3-generator, 9-bus power grid model problem

\[
\begin{align*}
T_{dqi} \frac{dE_{qi}'}{dt} &= -E_{qi}' - (X_{qi} - X_{qi}')I_{qi} \\
T_{dqi} \frac{dE_{qi}'}{dt} &= -E_{di}' - (X_{di} - X_{di}')I_{di} \\
\frac{d\delta_i}{dt} &= \omega - \omega_s \\
\frac{2H_i}{\omega_s} \frac{d\omega_i}{dt} &= T_{Mi} - E_{di}'I_{di} - (X_{qi}' - X_{di}')I_{di}I_{qi} - D_i(\omega_i - \omega_s) \\
T_{Ei} \frac{dE_{fdi}}{dt} &= -(K_{Ei} + S_{Ei}(E_{fdi}))E_{fdi} + V_{Ri} \\
T_{Fi} \frac{dR_{fi}}{dt} &= -R_{fi} + \frac{K_{fi}}{T_{fi}}E_{fdi} \\
T_{Ai} \frac{dV_{Ri}}{dt} &= -V_{Ri} + K_{Ai}R_{fi} - \frac{K_{Ai}K_{fi}}{T_{fi}}E_{fdi} + K_{Ai}(V_{refi} - V_i)
\end{align*}
\]

\[
\begin{align*}
0 &= E_{di}' - V_d - R_sI_{di} + X_{di}'I_{qi} \\
0 &= E_{qi}' - V_q - R_sI_{qi} - X_{qi}'I_{di} \\
\end{align*}
\]

The power grid inertia estimation problem

- Estimating the generator inertias during a dynamic transient generated by inducing a load disturbance
- Measurements: voltage phase and amplitude
- Forward and adjoint problems solved via the time-stepping (TS) component of PETSc that provides ODE and DAE integrators;
- The optimization problem is solved with the bounded limited memory quasi-Newton method for nonlinear minimization with bound constraints implemented in TAO.


Acknowledgment: 2015 and 2016 DOE Visiting Faculty program at Argonne National Laboratory
The MAP point (vs. the truth (23.64, 6.4, 3.01))

<table>
<thead>
<tr>
<th>$t_f$</th>
<th>$H_1$</th>
<th>$H_2$</th>
<th>$H_3$</th>
<th>#iter</th>
<th>$Err$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(a) $\Delta_t = 0.01$, $\Delta_{t,obs}^o = 0.05$</td>
</tr>
<tr>
<td>5.0</td>
<td>23.60</td>
<td>6.35</td>
<td>3.02</td>
<td>15</td>
<td>5.17e-03</td>
</tr>
<tr>
<td>3.0</td>
<td>23.79</td>
<td>6.39</td>
<td>3.06</td>
<td>9</td>
<td>1.01e-02</td>
</tr>
<tr>
<td>1.0</td>
<td>23.56</td>
<td>6.32</td>
<td>3.06</td>
<td>14</td>
<td>1.30e-02</td>
</tr>
<tr>
<td>0.8</td>
<td>23.67</td>
<td>6.54</td>
<td>2.95</td>
<td>11</td>
<td>1.76e-02</td>
</tr>
<tr>
<td>0.6</td>
<td>22.45</td>
<td>6.14</td>
<td>3.01</td>
<td>10</td>
<td>3.74e-02</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(b) $t_f = 1$, $\Delta_t = 0.01$</td>
</tr>
<tr>
<td>0.01</td>
<td>23.25</td>
<td>6.29</td>
<td>2.97</td>
<td>12</td>
<td>1.56e-02</td>
</tr>
<tr>
<td>0.02</td>
<td>23.81</td>
<td>6.50</td>
<td>3.00</td>
<td>12</td>
<td>9.79e-03</td>
</tr>
<tr>
<td>0.05</td>
<td>23.56</td>
<td>6.32</td>
<td>3.06</td>
<td>14</td>
<td>1.30e-02</td>
</tr>
<tr>
<td>0.10</td>
<td>22.91</td>
<td>6.42</td>
<td>3.06</td>
<td>13</td>
<td>1.11e-02</td>
</tr>
<tr>
<td>0.35</td>
<td>23.53</td>
<td>6.23</td>
<td>2.98</td>
<td>11</td>
<td>1.69e-02</td>
</tr>
</tbody>
</table>

We compare the MAP estimate to the truth using a relative error metric:

$$Err = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (m(i) - m_{true}(i))^2 / m_{true}(i)^2}.$$
The effect of the number of observation buses on the ability to recover the inertia parameter

<table>
<thead>
<tr>
<th>excluded buses</th>
<th>$H_1$</th>
<th>$H_2$</th>
<th>$H_3$</th>
<th>#iter</th>
</tr>
</thead>
<tbody>
<tr>
<td>none</td>
<td>23.56</td>
<td>6.32</td>
<td>3.06</td>
<td>14</td>
</tr>
<tr>
<td>1, 4</td>
<td>23.82</td>
<td>6.44</td>
<td>3.13</td>
<td>9</td>
</tr>
<tr>
<td>1, 4, 2, 7</td>
<td>23.86</td>
<td>6.56</td>
<td>3.03</td>
<td>10</td>
</tr>
<tr>
<td>all except 9</td>
<td>23.45</td>
<td>6.29</td>
<td>3.02</td>
<td>9</td>
</tr>
</tbody>
</table>

- The first column shows the excluded observation buses;
- The second, third and forth columns indicate the inverse solution, i.e., the MAP point;
- The fifth column (#iter) indicates the number of iterations for the optimization solver to converge.

3-generator 9-bus test case system.
The variance of the Gaussian(ized) posterior

Left: The trace of the Gaussianized posterior covariance. Right: A "whiskers boxplot" of the prior and posterior mean and variances for load and noise values (4.25, 0.01) and (7,0.1), respectively. The central mark is the median, the edges of the box are the 25'th and 75'th percentiles and the "whiskers" extend to the most extreme data points.
Summary and conclusions

- Quantifying the uncertainty in the inversions, leads to a large-scale Bayesian inverse problem governed by complex model.
- Hessian information is important for uncertainty quantification in inverse problems:
  - for posterior pdfs that are Gaussian, inverse of the Hessian is the covariance matrix;
  - for parameter-to-observable maps that are approximately linear over the range of likely parameters, the inverse of the Hessian approximates the covariance;
  - for infinite-dimensional parameters $m$, care is necessary to obtain proper discretizations of infinite-dimensional variables.
- Hessian manipulations can be made tractable by low-rank approximation of (prior-preconditioned) Hessian of the data misfit term.
- For the ice sheet flow inverse problem, we have seen that the computational work required for the Bayesian inversion—measured in number of forward (and adjoint) PDE solves—is independent of the state dimension and parameter dimension.
- Exploiting Hessian (of the log posterior) via a stochastic Newton method allows one to adapt MCMC proposal to local structure of posterior, leading to several orders of magnitude increase in convergence.
- For the power grid application, we learned that even with very little information we can produce a reasonable estimate for the inertial parameter.